LECTURE 4

Signal Detection & Statistics

Introduction

Signal detection is fundamental to astronomy

Instruments operate near sensitivity limits
Detections of individual photons
Faint sources are hard to distinguish from noise

Signal detection techniques depend on

- Wavelength
- Instrument → physical process of detection photons
- Source intensity → number of photons
- Prior knowledge of the source

Need to understand the statistical properties of the data

Astronomical Images

- Contain a hierarchy of sources
    - Star
    - Cluster
    - Galaxy
    - Group of Galaxies
- Contain sources with different shapes
    - Point sources
        - Stars
        - Nucleated galaxies
    - Extended sources
        - Galaxies
        - Nebulae
    - Blended sources
        - Sources too close to separate

Some Basic Statistical Concepts

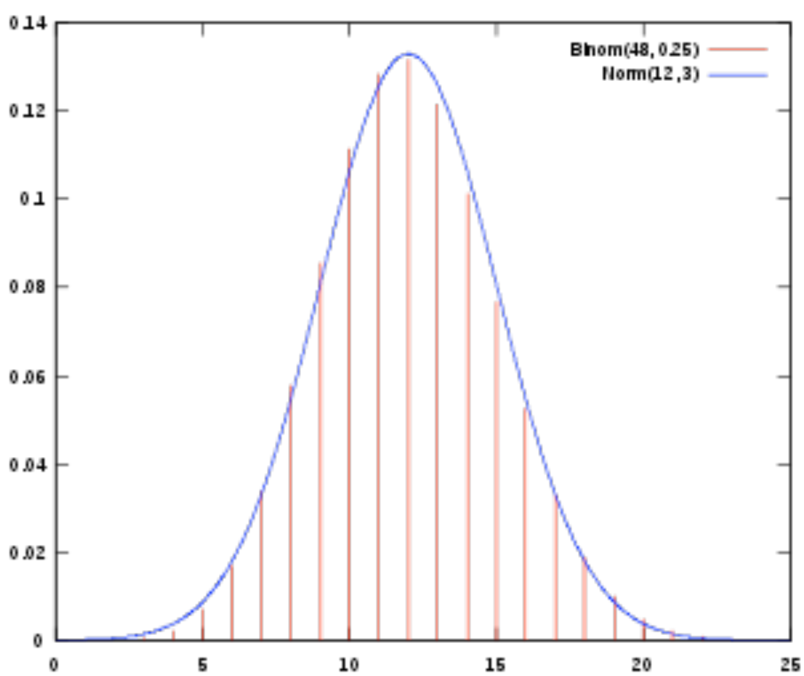
Mean

The **mean** is the sum of the data divided by the number of data points.

$$<x> = \sum_{i=1,N} x_i$$

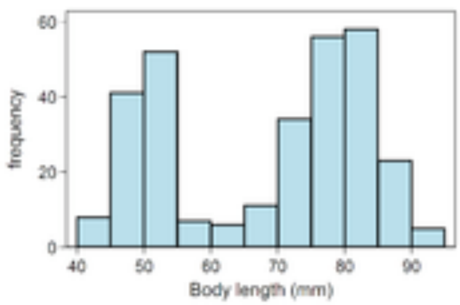
Advantages:

- Simple to compute
- Works well for symmetric distributions



Disadvantages:

- Sensitive to outliers
- Not good for multi-peaked distributions



Median

The **median** is the middle value of a sorted set of data. For example:

The data set {1,6,3,57,2,4,5} sorts to {1,2,3,**4**,5,6,57}. This middle value is 4, so the median is 4.

Advantages

- Insensitive to outliers

Disadvantages

- Computationally intensive

Mode

The **Mode** is the most common value in a data set. It is the peak of the probability distribution. For example:

The data set {1,2,3,4,2,3,4,3,4,4} has one 1, two 2s, three 3s, and four 4s. The most common value is 4, so the mode is 4.

Advantages:

- Not sensitive to outliers or the shape of the distribution
- Returns a typical value

Disadvantages:

- Not all data sets have a mode
  - What is the mode of {1,2,3,4,5}?
- Computationally intensive
- The mode can be significantly different from the mean or median for a highly skewed distribution

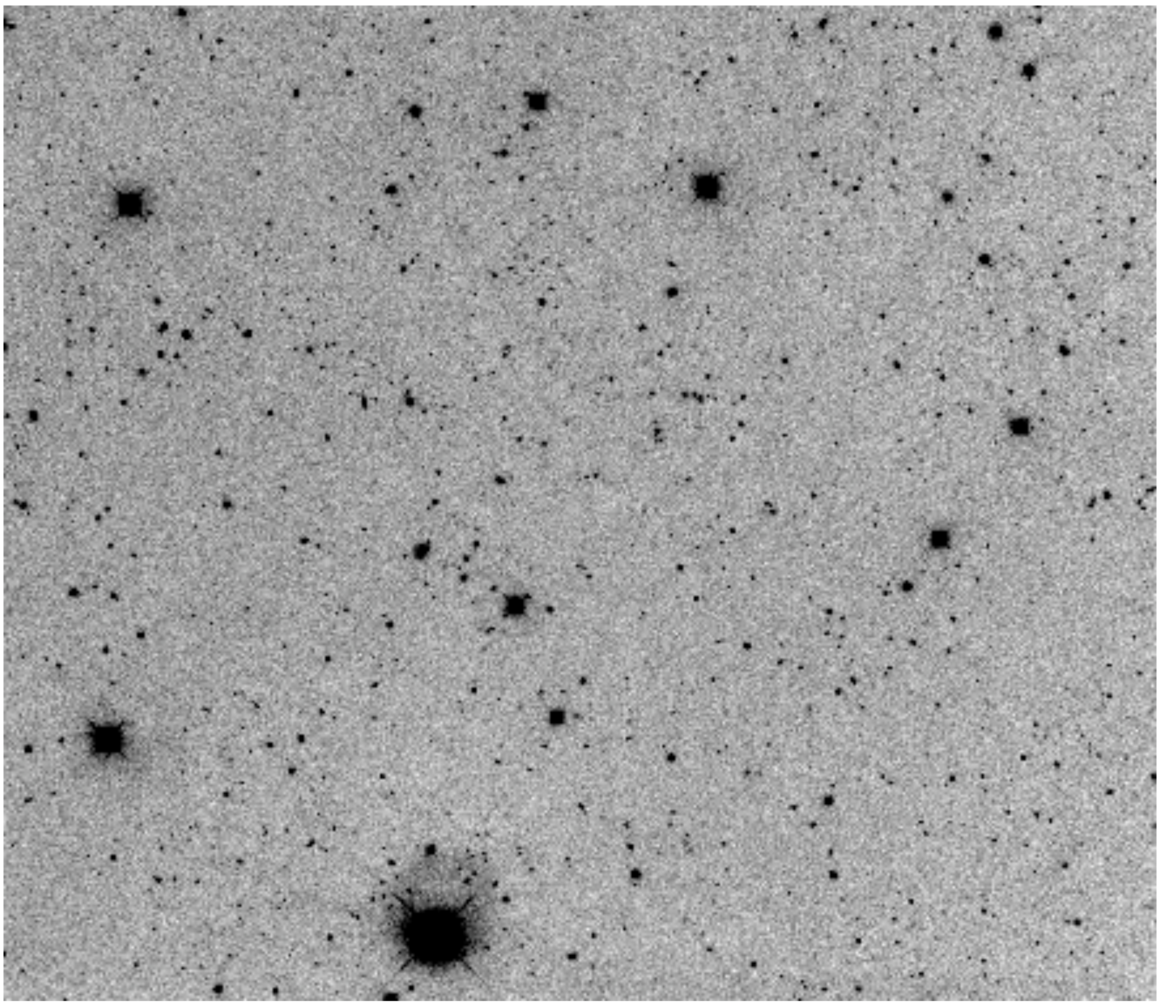
Standard Deviation

The **standard deviation** of a distribution is a measure of the variation in the data set.

$$\sigma^2 = 1 / (N-1) \sum_{i=1,N} (<x_i^2> - <x>^2)$$

Source Detection

GRB 050416A
Nordic Optical Telescope
DFOSC — 300 s — I band



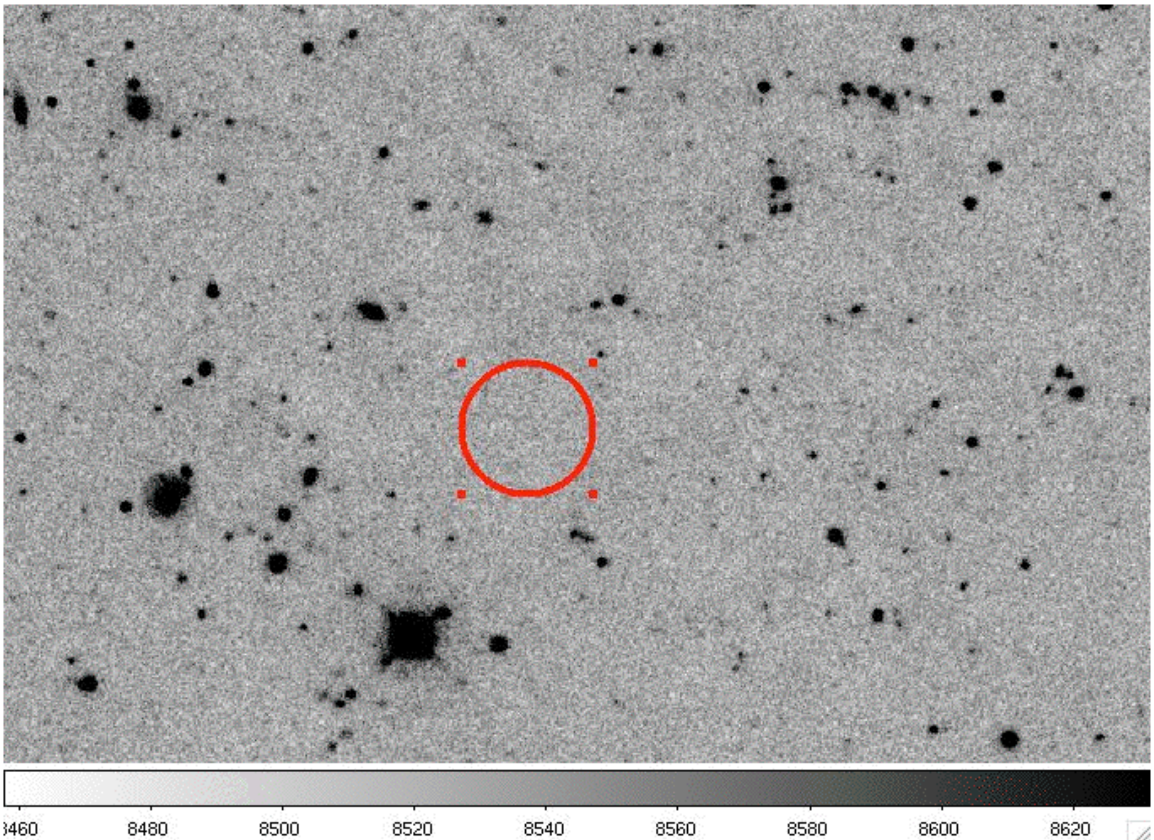
- Moderately crowded field
- No large background gradient
- More faint sources than bright ones
- Most sources are point sources

How to tell the sources from the noise?
- Easy by eye for bright sources
- Hard by eye for faint sources
- Need an automated process

Background Fluctuations
- Faint sources
- Noise

Sources of Noise
- Poisson noise
    - Intrinsic to photon counting statistics
    - Mean noise = sqrt(signal)
    - Observed counts = signal + noise
- Sky Noise
    - Poisson noise in the background
- Read-Out Noise
    - Noise from detector electronics
- Flat-Field Variance
    - Variance due to pixel-to-pixel variations in detector sensitivity
    - Typically about 1%–2%.

Measure the Background



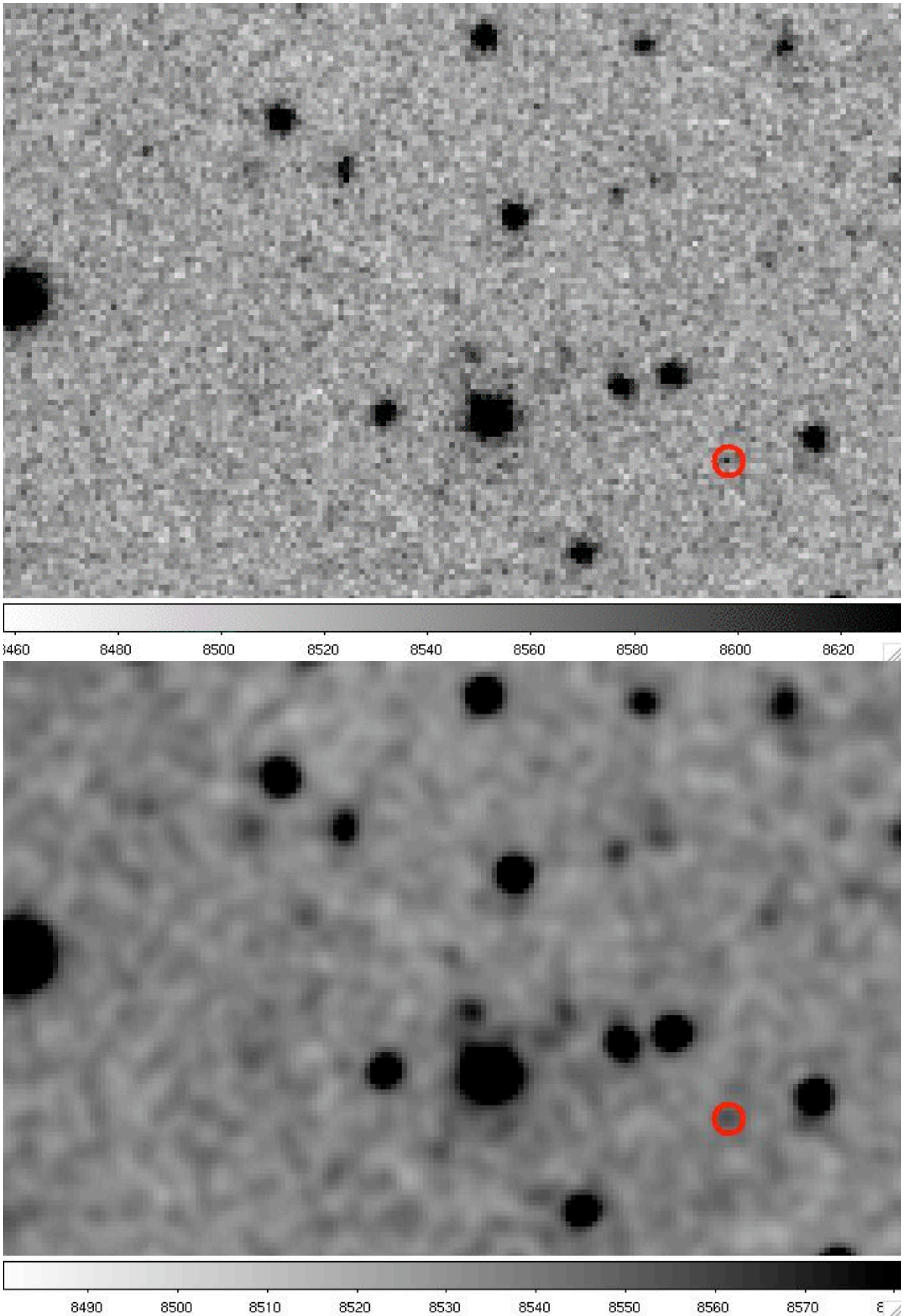
- Estimate the background
- Pick a region with no sources

- Pick a statistics
  - Mean
    - can be biased by faint, undetected sources
  - Median
    - usually reliable
    - computationally expensive
  - Mode
    - Usually reliable
    - Computational tricks exist
    - *Mode is often used*
- Background is usually determined locally
  - Compensates for variable background
- Measure the standard deviation in the sky.
- Automated methods

Filter the Image

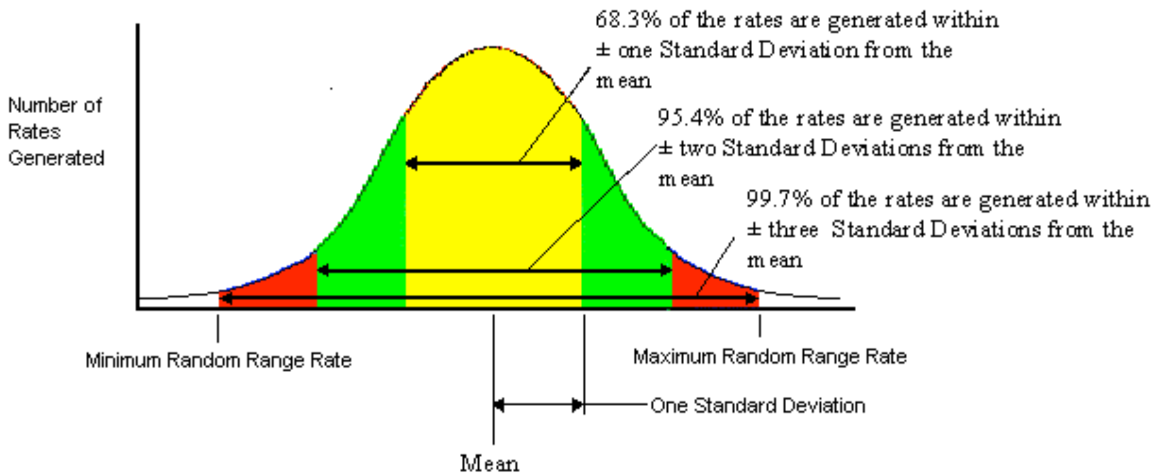
- Want to find astrophysical sources, not noise spikes
- Smooth the image with a matched filter to reduce noise.

- Atmospheric seeing, and the optics, cause photons from a point source, such as a star, to be spread over several pixels on a detector. This is called **the point-spread function**.
- In our example the width of the point-spread function is 1.2 arcseconds.
  - Angular size of Solar mass star at 10 pc: 0.001 arcsec
  - Diffraction limit: 0.08 arcsec
  - Pixel scale: 0.39 arcsec
- CCD point-spread functions are roughly Gaussian (to a first approximation)
- Pick a filter that matches the point sources in the image.
- Smooth the image
- Any source smaller than the point-spread function will be smoothed over.
- Point sources and extended sources will still be present.


8480 8500 8520 8540 8560 8580 8600 8620
8490 8500 8510 8520 8530 8540 8550 8560 8570

Detect Sources

- Scan the smoothed image and look for pixels that are > *N*-sigma brighter than the background.
- *N* is the detection threshold



- For a Gaussian distribution of noise (typical for a CCD)

***N*-sigma**	**Chance of False Detection**
1	31.7%
2	4.6%
3	0.3%

- One way to further reduce the probability of a chance detection is to require that there be at least M adjacent pixels that are above the detection threshold for each source.

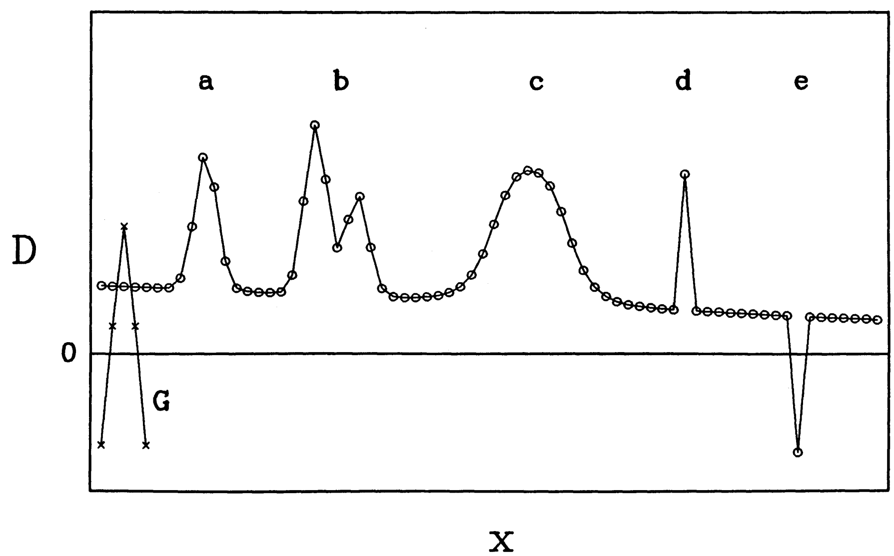
- For M = 5 this would be a detection

	O	O	O	
		X	O	
		O		

- And this would not.

	O			
		X	O	
		O		

Example of Detected Sources



a. This is a star.
b. This is two stars that are blended. The point-spread functions of the two stars overlap.
c. This is a galaxy that is slightly larger than the point-spread function. It is only barely resolved.
d. This is either a noise spike, or a cosmic ray.
e. This is a bad (cold) pixel.